RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/751,235
Source: IFWO
Date Processed by STIC: 11-18-04

ENTERED

IFWO

RAW SEQUENCE LISTING   DATE: 11/18/2004
PATENT APPLICATION: US/10/751,235   TIME: 12:20:26

Input Set : A:\08604~2.txt
Output Set: N:\CRF4\11182004\J751235.raw

```
 3 <110> APPLICANT: DellaPenna, Dean
 4       Tian, Li
 5       Kim, Joonyul
 7 <120> TITLE OF INVENTION: Novel Carotenoid Hydroxylases for Use in Engineering
Carotenoid
 8       Metabolism in Plants
10 <130> FILE REFERENCE: MSU-08604
12 <140> CURRENT APPLICATION NUMBER: 10/751,235
13 <141> CURRENT FILING DATE: 2004-01-02
15 <160> NUMBER OF SEQ ID NOS: 74
17 <170> SOFTWARE: PatentIn version 3.2
19 <210> SEQ ID NO: 1
20 <211> LENGTH: 77
21 <212> TYPE: PRT
22 <213> ORGANISM: Arabidopsis thaliana
24 <400> SEQUENCE: 1
26 Leu Gln Pro Tyr Ala Glu Asp Gly Ser Ala Val Asn Met Glu Ala Lys
27 1               5                   10                  15
30 Phe Ser Gln Met Thr Leu Asp Val Ile Gly Leu Ser Leu Phe Asn Tyr
31              20                  25                  30
34 Asn Phe Asp Ser Leu Thr Thr Asp Ser Pro Val Ile Glu Ala Val Tyr
35           35                  40                  45
38 Thr Ala Leu Lys Glu Ala Glu Leu Arg Ser Thr Asp Leu Leu Pro Tyr
39     50                  55                  60
42 Trp Lys Ile Asp Ala Leu Cys Lys Ile Val Pro Arg Gln
43 65                  70                  75
46 <210> SEQ ID NO: 2
47 <211> LENGTH: 77
48 <212> TYPE: PRT
49 <213> ORGANISM: Arabidopsis thaliana
51 <400> SEQUENCE: 2
53 Leu Asp Ala Ala Ala Leu Lys Gly Glu Glu Val Glu Met Glu Ser Leu
54 1               5                   10                  15
57 Phe Ser Arg Leu Thr Leu Asp Ile Ile Gly Lys Ala Val Phe Asn Tyr
58              20                  25                  30
61 Asp Phe Asp Ser Leu Thr Asn Asp Thr Gly Val Ile Glu Ala Val Tyr
62           35                  40                  45
65 Thr Val Leu Arg Glu Ala Glu Asp Arg Ser Val Ser Pro Ile Pro Val
66     50                  55                  60
69 Trp Asp Ile Pro Ile Trp Lys Asp Ile Ser Pro Arg Gln
70 65                  70                  75
73 <210> SEQ ID NO: 3
74 <211> LENGTH: 84
75 <212> TYPE: PRT
```

```
 76 <213> ORGANISM: Arabidopsis thaliana
 78 <400> SEQUENCE: 3
 80 Glu Lys Leu Ile Arg Glu Lys Glu Thr Ser Ser Gly Glu Asp Thr Ile
 81 1               5                   10                  15
 84 Glu Leu Asp Leu Glu Ala Glu Phe Ser Ser Leu Ala Leu Asp Ile Ile
 85          20                  25                  30
 88 Gly Leu Ser Val Phe Asn Tyr Asp Phe Gly Ser Val Thr Lys Glu Ser
 89          35                  40                  45
 92 Pro Val Ile Lys Ala Val Tyr Gly Thr Leu Phe Glu Ala Glu His Arg
 93       50                  55                  60
 96 Ser Thr Phe Tyr Phe Pro Tyr Trp Asn Phe Pro Pro Ala Arg Trp Ile
 97 65                  70                  75                  80
100 Val Pro Arg Gln
104 <210> SEQ ID NO: 4
105 <211> LENGTH: 539
106 <212> TYPE: PRT
107 <213> ORGANISM: Arabidopsis thaliana
109 <400> SEQUENCE: 4
111 Met Glu Ser Ser Leu Phe Ser Pro Ser Ser Ser Ser Tyr Ser Ser Leu
112 1               5                   10                  15
115 Phe Thr Ala Lys Pro Thr Arg Leu Leu Ser Pro Lys Pro Lys Phe Thr
116              20                  25                  30
119 Phe Ser Ile Arg Ser Ser Ile Glu Lys Pro Lys Pro Lys Leu Glu Thr
120          35                  40                  45
123 Asn Ser Ser Lys Ser Gln Ser Trp Val Ser Pro Asp Trp Leu Thr Thr
124       50                  55                  60
127 Leu Thr Arg Thr Leu Ser Ser Gly Lys Asn Asp Glu Ser Gly Ile Pro
128 65                  70                  75                  80
131 Ile Ala Asn Ala Lys Leu Asp Asp Val Ala Asp Leu Leu Gly Gly Ala
132              85                  90                  95
135 Leu Phe Leu Pro Leu Tyr Lys Trp Met Asn Glu Tyr Gly Pro Ile Tyr
136              100                 105                 110
139 Arg Leu Ala Ala Gly Pro Arg Asn Phe Val Ile Val Ser Asp Pro Ala
140          115                 120                 125
143 Ile Ala Lys His Val Leu Arg Asn Tyr Pro Lys Tyr Ala Lys Gly Leu
144      130                 135                 140
147 Val Ala Glu Val Ser Glu Phe Leu Phe Gly Ser Gly Phe Ala Ile Ala
148 145                 150                 155                 160
151 Glu Gly Pro Leu Trp Thr Ala Arg Arg Arg Ala Val Val Pro Ser Leu
152              165                 170                 175
155 His Arg Arg Tyr Leu Ser Val Ile Val Glu Arg Val Phe Cys Lys Cys
156          180                 185                 190
159 Ala Glu Arg Leu Val Glu Lys Leu Gln Pro Tyr Ala Glu Asp Gly Ser
160          195                 200                 205
163 Ala Val Asn Met Glu Ala Lys Phe Ser Gln Met Thr Leu Asp Val Ile
164      210                 215                 220
167 Gly Leu Ser Leu Phe Asn Tyr Asn Phe Asp Ser Leu Thr Thr Asp Ser
168 225                 230                 235                 240
171 Pro Val Ile Glu Ala Val Tyr Thr Ala Leu Lys Glu Ala Glu Leu Arg
```

```
172                  245                250                  255
175 Ser Thr Asp Leu Leu Pro Tyr Trp Lys Ile Asp Ala Leu Cys Lys Ile
176              260              265              270
179 Val Pro Arg Gln Val Lys Ala Glu Lys Ala Val Thr Leu Ile Arg Glu
180          275            280              285
183 Thr Val Glu Asp Leu Ile Ala Lys Cys Lys Glu Ile Val Glu Arg Glu
184      290            295              300
187 Gly Glu Arg Ile Asn Asp Glu Glu Tyr Val Asn Asp Ala Asp Pro Ser
188 305            310              315                  320
191 Ile Leu Arg Phe Leu Leu Ala Ser Arg Glu Glu Val Ser Ser Val Gln
192          325              330                  335
195 Leu Arg Asp Asp Leu Leu Ser Met Leu Val Ala Gly His Glu Thr Thr
196          340              345              350
199 Gly Ser Val Leu Thr Trp Thr Leu Tyr Leu Leu Ser Lys Asn Ser Ser
200          355              360              365
203 Ala Leu Arg Lys Ala Gln Glu Glu Val Asp Arg Val Leu Glu Gly Arg
204      370            375              380
207 Asn Pro Ala Phe Glu Asp Ile Lys Glu Leu Lys Tyr Ile Thr Arg Cys
208 385            390              395                  400
211 Ile Asn Glu Ser Met Arg Leu Tyr Pro His Pro Pro Val Leu Ile Arg
212          405              410                  415
215 Arg Ala Gln Val Pro Asp Ile Leu Pro Gly Asn Tyr Lys Val Asn Thr
216          420              425              430
219 Gly Gln Asp Ile Met Ile Ser Val Tyr Asn Ile His Arg Ser Ser Glu
220          435              440              445
223 Val Trp Glu Lys Ala Glu Glu Phe Leu Pro Glu Arg Phe Asp Ile Asp
224      450            455              460
227 Gly Ala Ile Pro Asn Glu Thr Asn Thr Asp Phe Lys Phe Ile Pro Phe
228 465            470              475                  480
231 Ser Gly Gly Pro Arg Lys Cys Val Gly Asp Gln Phe Ala Leu Met Glu
232          485              490                  495
235 Ala Ile Val Ala Leu Ala Val Phe Leu Gln Arg Leu Asn Val Glu Leu
236          500              505              510
239 Val Pro Asp Gln Thr Ile Ser Met Thr Thr Gly Ala Thr Ile His Thr
240      515            520              525
243 Thr Asn Gly Leu Tyr Met Lys Val Ser Gln Arg
244      530            535
247 <210> SEQ ID NO: 5
248 <211> LENGTH: 2467
249 <212> TYPE: DNA
250 <213> ORGANISM: Arabidopsis thaliana
252 <400> SEQUENCE: 5
253 atggagtctt cactcttttc tccatcttcc tcttcttact cttctctctt cactgcaaaa    60
255 cctacgcgtc ttttatcacc aaaacccaaa ttcacattct ccatcagatc ctccattgag   120
257 aaacccaaac ccaaactcga gaccaattca tcgaaatccc aatcatgggt cagtcccgat   180
259 tggctcacaa cactcactcg tacccttttcc tcaggaaaaa acgacgagtc aggtatacca   240
261 atcgcgaacg cgaagctcga cgatgtcgct gatctcctcg gaggtgctct cttcttacct   300
263 ctctacaaat ggatgaatga gtacggaccc atttaccgtc tcgctgctgg tcctcgtaat   360
265 ttcgtaattg tgagcgaccc agcgatagct aaacatgttt tgaggaatta tccaaagtac   420
```

```
267 gctaaaggct tagtcgctga agtctctgaa tttctatttg gttcgggttt cgctatcgct    480
269 gaaggacctc tttggacagt aatttcatct cctcctatct caattttgaa gttttttggaa   540
271 ttgtggaagt aatgtgtgac tgtcttgtat gataagtaac tctaatttta gggtttagat    600
273 tccaatcttc tctattgggc ttagctgaag tctgattttt tacataggcg aggcgtagag    660
275 cggtggttcc atcgcttcac aggaggtatt tgtctgtgat tgtggagaga gtattctgca    720
277 aatgtgcaga gaggcttgtt gagaagttgc agccttatgc agaagacgga agtgctgtga    780
279 atatggaagc gaagttctct cagatgacac ttgatgtcat tgggttgtct cttttttaact  840
281 acaatttcga ttctttgact actgatagtc ctgtcattga agctgtttac actgctctta    900
283 aagaagctga gcttcgttct actgatcttc tgccatattg gaaggcaagt ttcctgtgtt    960
285 ttttctgtgg tttgttgatt gtgtggaaca attggattct tgttaattga gagggtttgg    1020
287 ttgttttttt cagatcgatg cattgtgtaa gatagtcccg agacaggtga aagctgaaaa     1080
289 ggctgtaact ttgataaggg aaactgttga agaccttatt gctaagtgta aagaaattgt    1140
291 cgaaagagaa ggcgaaagaa tcaatgatga ggagtatgta aatgatgctg acccaagtat    1200
293 cctgcgtttc ttgcttgcaa gcagagaaga ggtttaaact ttttttcctta agtttataag   1260
295 caaatttggc ctttcattat cgcataatcg aagctgatgt tgcattgtga gggttttcag    1320
297 gtatcaagtg tgcagttacg ggatgatctt ctctcaatgc tcgtagcggg tcatgaaacc    1380
299 actggatctg tcctcacttg gacactttat ctcctaagta aggtacctta atgtatcttc    1440
301 tactttgcta tgctagagaa tttacttgga tgggagcttc tctgttctca tttacctctt    1500
303 caaattctct atgttcatag aactcatctg cattaaggaa agcacaagaa gaagtagaca    1560
305 gagtgttaga aggaagaaac ccggctttcg aggatataaa ggagttgaag tacatcactc    1620
307 gttgtataaa cgagtcaatg cgtctctatc ctcatcctcc tgtaagcaat caagctcatc    1680
309 tctctaatta ttcatgaact aaattttctg attgatttgt ttcctggtag gtcttgataa   1740
311 gaagagctca agttcctgac attcttcctg ggaactataa ggtcaatacc ggacaagaca    1800
313 ttatgatttc agtctataac atccatcgtt cttccgaggt acagttctct tccttctctc    1860
315 gtccatagta taacataggg gagcctaatc cttctcttca atgatctttg tgtggttcgg    1920
317 atatctaacc ggagtggaca ttcctagtat tacattcatg cccacatttc ttatgtgttt    1980
319 gttgtttgtt attccaaagg tatgggaaaa agctgaggaa tttctgcctg aacgattcga    2040
321 catagatggc gcaatcccta acgaaacaaa cactgatttc aagtaaactc agtagaacac    2100
323 atcttttgac acaaactact gaatcaagat tagtggtttt gattagggaa tttaaaagat    2160
325 gattttcttt tttcaccaga ttcatcccat tcagtggagg gcctagaaaa tgtgtaggcg    2220
327 atcagtttgc attgatggag gcaattgtgg cactcgcggt gtttcttcag cggttaaacg    2280
329 ttgagctggt tcctgatcag accattagca tgaccacagg agcaaccata cacaccacca    2340
331 atgtatgcca atgttctcac actcgagaga ttaatgagag tgtctgtttt gtttagaatg    2400
333 attccaattt ctctaatgct gatattttca atttcaggga ttgtatatga aggtgagcca    2460
335 aaggtaa                                                                2467
338 <210> SEQ ID NO: 6
339 <211> LENGTH: 4170
340 <212> TYPE: DNA
341 <213> ORGANISM: Arabidopsis thaliana
343 <400> SEQUENCE: 6
344 cctggtcaag agagagcaga aattgctgat acgcatcttt gtcagagaga agcttcctga      60
346 gttcatccac gctacagaaa agtatacact attcaaaaag aagacataca aaaaaaacgc     120
348 acactcgaga gaaggataac aaacaaacaa acaaaaagag caaacctttt gtctttcaag     180
350 aagacgataa tgccagctgc ttcacccggt gatacttgtg ctgaaatctg acctgaactt     240
352 tgaggccggg atgagcttgg cgagctaact aaagaaggcg agtaccatgg cgactgtgaa     300
354 gaagcttcct gaggacgaga ttgcccttgt tgttggtctt ttgatcccct tcacgtacac     360
356 aacaaaaaaa agaaacatca tcaactaact atctcagctt ccaagttaat ataatacatc     420
358 acctcataaa cttgaaaggt gcatttatac agcacatcat tgtaaaccta atatcagtaa     480
360 ctatgaaccc taaatcagta gctgagcaaa atctacactt gtcaattcac ctcaaaaacc     540
```

```
RAW SEQUENCE LISTING              DATE: 11/18/2004
PATENT APPLICATION:  US/10/751,235   TIME: 12:20:26

Input Set : A:\08604~2.txt
Output Set: N:\CRF4\11182004\J751235.raw

362 tcaaatatat cattatagct caacaaacta cacaacaaca agatttcctt aatcgctaac    600
364 agacctaatg tctggatcac gcaatttcat caataacata tgaatcagat ggcaaaaaat    660
366 tcgcaaaatc ataccagaaa ttgaacatcg tttctagaat tcgccgacta acgaaaggaa    720
368 ttgacttttt cttgttctct tgtacttctt cttcctacca gatcgacaaa atttaacaac    780
370 tttacagaaa cagagtttac agagtttcaa agaaaattcg atcttctctg tttcgtttca    840
372 gattttgcta cttgctttta aaaggcccaa atacgaggag gcccatatta agttttgatt    900
374 aacaaaacgt agtcgtttta tgcatcgtcg tccttatcca cgcgtaacgg ttcgcgtttc    960
376 tacacagagt caaatatttc tgcacggaag cttcgaaaag aggtcatcaa tggagtcttc   1020
378 actcttttct ccatcttcct cttcttactc ttctctcttc actgcaaaac ctacgcgtct   1080
380 tttatcacca aaacccaaat tcacattctc catcagatcc tccattgaga aacccaaacc   1140
382 caaactcgag accaattcat cgaaatccca atcatgggtc agtcccgatt ggctcacaac   1200
384 actcactcgt accctttcct caggaaaaaa cgacgagtca ggtataccaa tcgcgaacgc   1260
386 gaagctcgac gatgtcgctg atctcctcgg aggtgctctc ttcttacctc tctacaaatg   1320
388 gatgaatgag tacggaccca tttaccgtct cgctgctggt cctcgtaatt cgtaattgt    1380
390 gagcgaccca gcgatagcta aacatgtttt gaggaattat ccaaagtacg ctaaaggctt   1440
392 agtcgctgaa gtctctgaat ttctatttgg ttcgggtttc gctatcgctg aaggacctct   1500
394 ttggacagta atttcatctc ctcctatctc aattttgaag tttttggaat tgtggaagta   1560
396 atgtgtgact gtcttgtatg ataagtaact ctaattttag ggtttagatt ccaatcttct   1620
398 ctattgggct tagctgaagt ctgatttttt acataggcga ggcgtagagc ggtggttcca   1680
400 tcgcttcaca ggaggtattt gtctgtgatt gtggagagag tattctgcaa atgtgcagag   1740
402 aggcttgttg agaagttgca gccttatgca gaagacggaa gtgctgtgaa tatggaagcg   1800
404 aagttctctc agatgacact tgatgtcatt gggttgtctc tttttaacta caatttcgat   1860
406 tctttgacta ctgatagtcc tgtcattgaa gctgtttaca ctgctcttaa agaagctgag   1920
408 cttcgttcta ctgatcttct gccatattgg aaggcaagtt tcctgtgttt tttctgtggt   1980
410 ttgttgattg tgtggaacaa ttggattctt gttaattgag agggtttggt tgtttttttc   2040
412 agatcgatgc attgtgtaag atagtcccga gacaggtgaa agctgaaaag gctgtaactt   2100
414 tgataaggga aactgttgaa gaccttattg ctaagtgtaa agaaattgtc gaaagagaag   2160
416 gcgaaagaat caatgatgag gagtatgtaa atgatgctga cccaagtatc ctgcgtttct   2220
418 tgcttgcaag cagagaagag gtttaaactt tttttcttaa gtttataagc aaatttggcc   2280
420 tttcattatc gcataatcga agctgatgtt gcattgtgag ggttttcagg tatcaagtgt   2340
422 gcagttacgg gatgatcttc tctcaatgct cgtagcgggt catgaaacca ctggatctgt   2400
424 cctcacttgg acactttatc tcctaagtaa ggtaccttaa tgtatcttct actttgctat   2460
426 gctagagaat ttacttggat gggagcttct ctgttctcat ttacctcttc aaattctcta   2520
428 tgttcataga actcatctgc attaaggaaa gcacaagaag aagtagacag agtgttagaa   2580
430 ggaagaaacc cggctttcga ggatataaag gagttgaagt acatcactcg ttgtataaac   2640
432 gagtcaatgc gtctctatcc tcatcctcct gtaagcaatc aagctcatct ctctaattat   2700
434 tcatgaacta aattttctga ttgatttgtt tcctggtagg tcttgataag aagagctcaa   2760
436 gttcctgaca ttcttcctgg gaactataag gtcaataccg gacaagacat tatgatttca   2820
438 gtctataaca tccatcgttc ttccgaggta cagttctctt ccttctctcg tccatagtat   2880
440 aacatagggg agcctaatcc ttctcttcaa tgatctttgt gtggttcgga tatctaaccg   2940
442 gagtggacat tcctagtatt acattcatgc ccacatttct tatgtgtttg ttgtttgtta   3000
444 ttccaaaggt atgggaaaaa gctgaggaat ttctgcctga acgattcgac atagatggcg   3060
446 caatccctaa cgaaacaaac actgatttca agtaaactca gtagaacaca tcttttgaca   3120
448 caaactactg aatcaagatt agtggttttg attagggaat ttaaagatg attttctttt   3180
450 ttcaccagat tcatcccatt cagtggaggg cctagaaaat gtgtaggcga tcagtttgca   3240
452 ttgatggagg caattgtggc actcgcggtg tttcttcagc ggttaaacgt tgagctggtt   3300
454 cctgatcaga ccattagcat gaccacagga gcaaccatac acaccaccaa tgtatgccaa   3360
456 tgttctcaca ctcgagagat taatgagagt gtctgttttg tttagaatga ttccaatttc   3420
458 tctaatgctg atattttcaa tttcagggat tgtatatgaa ggtgagccaa aggtaaaaac   3480
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 11/18/2004
PATENT APPLICATION: US/10/751,235   TIME: 12:20:27

Input Set : **A:\08604~2.txt**
Output Set: **N:\CRF4\11182004\J751235.raw**

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:12; Xaa Pos. 1,3,4,6
Seq#:14; Xaa Pos. 2,3,5,6,7,9
Seq#:17; Xaa Pos. 529
Seq#:23; N Pos. 1587

VERIFICATION SUMMARY   DATE: 11/18/2004
PATENT APPLICATION: US/10/751,235   TIME: 12:20:27

Input Set : A:\08604~2.txt
Output Set: N:\CRF4\11182004\J751235.raw

L:579 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (8) SEQUENCE:
L:584 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (9) SEQUENCE:
L:657 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12 after pos.:0
L:701 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:14 after pos.:0
L:1012 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:17 after pos.:528
L:1441 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:23 after pos.:1560
L:4278 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (58) SEQUENCE: